CA Water Board Data Management Strategy and Open Data Initiative

•••

Greg Gearheart, Deputy Director
Office of Information Management and Analysis

For WQCC - October 26, 2017

Open Data Culture

Office of Information Management and Analysis (OIMA)

SWAMP Unit

Lori Webber (Sr ES)

Surface Water Ambient Monitoring Program - Overall Program Management and Support

SWAMP IQ Unit

Melissa Morris (Sr ES)

Quality Assurance and Data Management Support for SWAMP and the CA Env Data Exchange Network (CEDEN)

Data Integration Unit

Jarma Bennett (Sr Eng)

Data integration and analysis, CEDEN and CIWQS Project Management, “Open Data” Initiative

Additional OIMA Functions and Staff (cont.)

- Citizen Monitoring, Open Science and SWAMP Southern CA Support - Erickson Burres (ES Spec)
- Chief Data Scientist, AB1755 Lead and Water Economist - Rafael Maestu (RPS2)
- Performance Management and Watershed Outcomes - Bev Anderson-Abbs (ES Spec)
- Water Quality Monitoring Council Director - vacant (EPM I)
- Quality Officer, Quality Management Program Manager - Renee Spears (ES Spec)
- SWAMP Assessment and Reporting Lead Scientist (pesticides, CECs and others) - Dawit Tadesse (ES Spec)

Data Management Strategy (2017 update)

Annual Civic Engagement Events

- Data Fair (open house)
- Data Innovation Challenge (hackathons)
- Water / Data Science Synthesis
- Brown Bag Series of Speakers
- Other partnerships

Our Data Management Strategy Framework

- Based on Principles
- Will guide:
  - Data driven management
  - “Water” decisions
  - “Technology” decisions
  - Quality program
- Lists our data management values
- Encourages “data literacy”

Databases and Datasets at the CA Water Boards

- Over 20 enterprise database applications
- Water quality, water rights, drinking water, etc.
- Program data (e.g, facilities, activities) and environmental / ambient data (e.g., surface water and groundwater quality, water use, water conservation, etc.)
- 18 data resources on data.ca.gov → more all the time

Why do we collect, use and produce data?

- To inform our **data-driven management** and planning activities – performance report cards, workplans, resource assignment/augmentation, evaluating program effectiveness, and many others examples;
- To inform our **critical decisions** regarding our mission(s) and water management responsibilities – water allocation and use, water quality planning and "policies," permitting, program prioritization, and many other examples; and
- To provide **transparency** to our many partners and stakeholders for their use, interests and purposes.

Five Guiding Principles for WB Data Management

1. **Accessible ("Open first")**: our organization values transparency and strives to make all critical data available in machine readable datasets with metadata and data dictionaries
2. **Quality and integrity**: our data is of known and acceptable quality and we deploy practices to protect its integrity with standards and protocols
3. **Data literacy**: our whole organization understands its data needs and responsibilities, can speak the language of data science the staff and managers have robust data science capacity

Five Guiding Principles for WB Data Management (cont.)

4. **Use data to govern:** our organization uses data to govern, or makes decisions that are in the best interest of our mission(s)
5. **Govern our data:** our organization takes proactive steps to develop effective data and information technology management practices to ensure our data flows to where it is needed in a timely manner while complying with our data sharing policies

L1 - Data Collection

CIWQS Data

CIWQS Data Collection

Standards

Data Checkers

Data Checkers

Data

SWAMP Data Collection

Standards

L2 - Data Storage

CIWQS Database

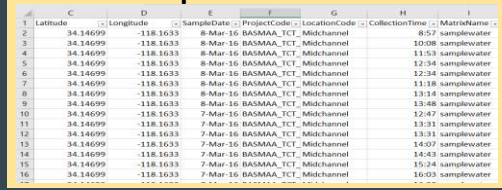
SWAMP / CEDEN Database

Other Databases

Web Services

Web Services

L3 - Open Data



CalData data.ca.gov Portal

CIWQS

SWAMP / CEDEN

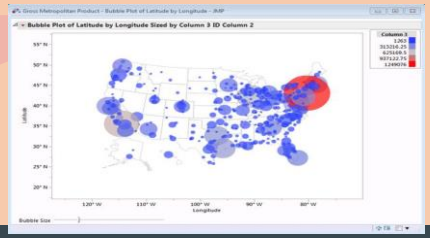
Other Open Data

Federal

CA DWR

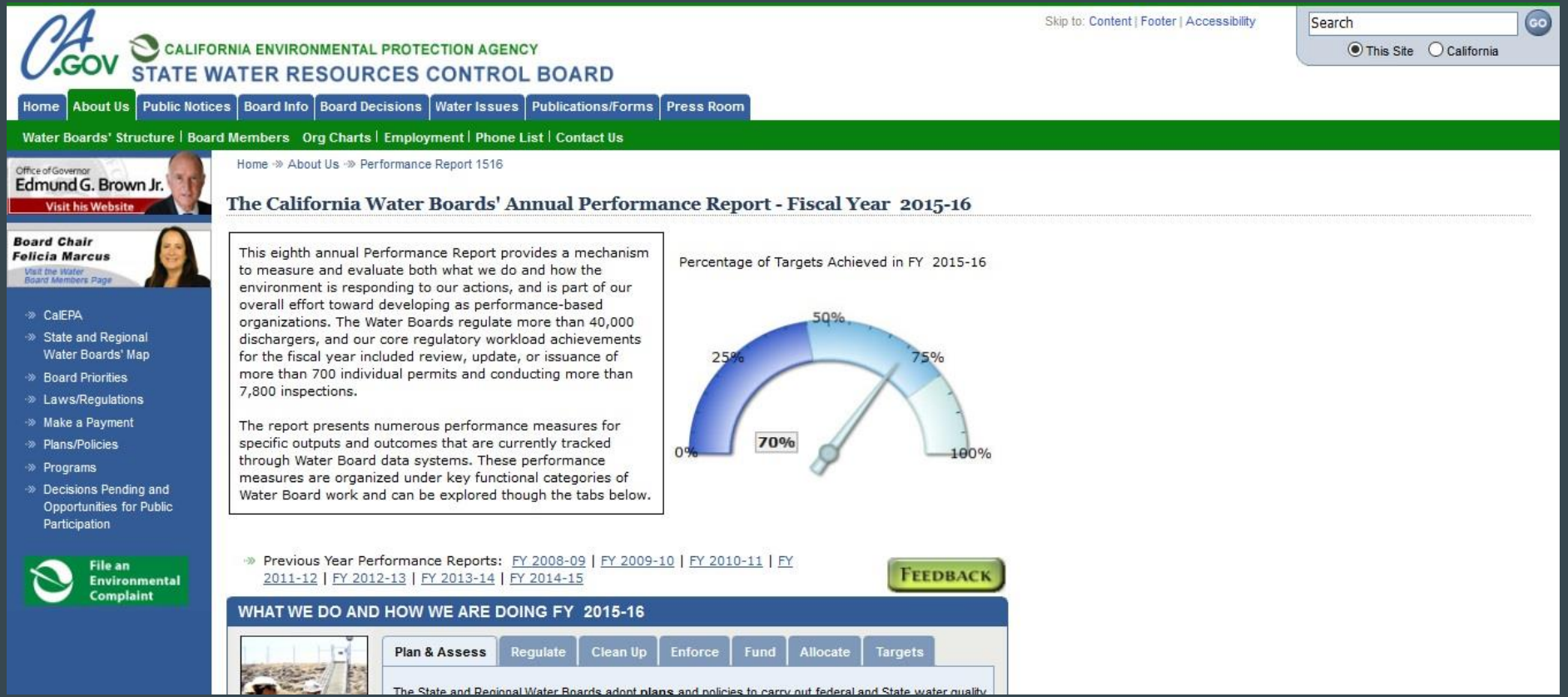
APIs / Web Services

L4 - Data Viz / Analysis



We run the Water Boards' performance program, made up of over 190 report cards on outputs covering both our program resources and expectations as well as water quality outcome stories.

Data Driven Management and Storytelling



https://www.waterboards.ca.gov/about_us/performance_report_1516/index.shtml

- CalEPA
- State and Regional Water Boards' Map
- Board Priorities
- Laws/Regulations
- Make a Payment
- Plans/Policies
- Programs
- Decisions Pending and Opportunities for Public Participation

File an Environmental Complaint

Home » Resources » Data Databases

SWAMP
Surface Water Ambient Monitoring Program

2017 Water Quality Status Report

SWAMP
Surface Water Ambient Monitoring Program

The California Water Boards' Water Data Center is proud to present the CA Water Quality Status Report. This report is an annual data-driven snapshot of the Water Board's water quality and environmental data. This inaugural version of the report is based solely on the surface water datasets available via the Surface Water Ambient Monitoring Program (SWAMP) and in future years we hope to expand this to include the groundwater, drinking water and water resource datasets available in our state. Our goal is to use data to inform both data storytelling (as in this inaugural report) and water quality indicators, including watershed report cards.

The 2017 Water Quality Status Report is organized around seven major themes that our team thought both individually and collectively tell important stories about the overall health of our state's surface waters. Each theme-specific story includes a brief background, a data analysis summary, an overview of management actions, and access to the raw data. All of the data in the stories is available at the State of California's Open Data Portal, https://data.ca.gov/dataset/2017-california-water-quality-status-report.

For more information, please contact the Office of Information Management and Analysis (OIMA) OIMA Help Desk.

Contaminants and Toxicity in Stream Sediments

Bioaccumulation of Contaminants in Fish

Setting Flow Targets to Support Biological Integrity in SoCal Streams

https://www.waterboards.ca.gov/resources/data_databases/wq_status_report.shtml

First Annual “Water Quality Status Report” - Series of Data Stories, Piloting of Tool for Outcomes

Water Boards

Setting Flow Targets to Support Biological Integrity in Southern California Streams

Among the range of approaches available for setting flow targets that support biological integrity, a recently completed project in southern California utilized the Ecological Limits of Hydrologic Alteration (ELOHA) framework to assess the effect of flow alteration on the condition of benthic macroinvertebrate (BMI) communities across the region. The framework establishes recommended flow targets using a process that includes estimation of flow alteration and development of flow-ecology relationships based on the response of biological communities to changes in flow.

Biological Data Provide Ecological Context for Hydrologic Data

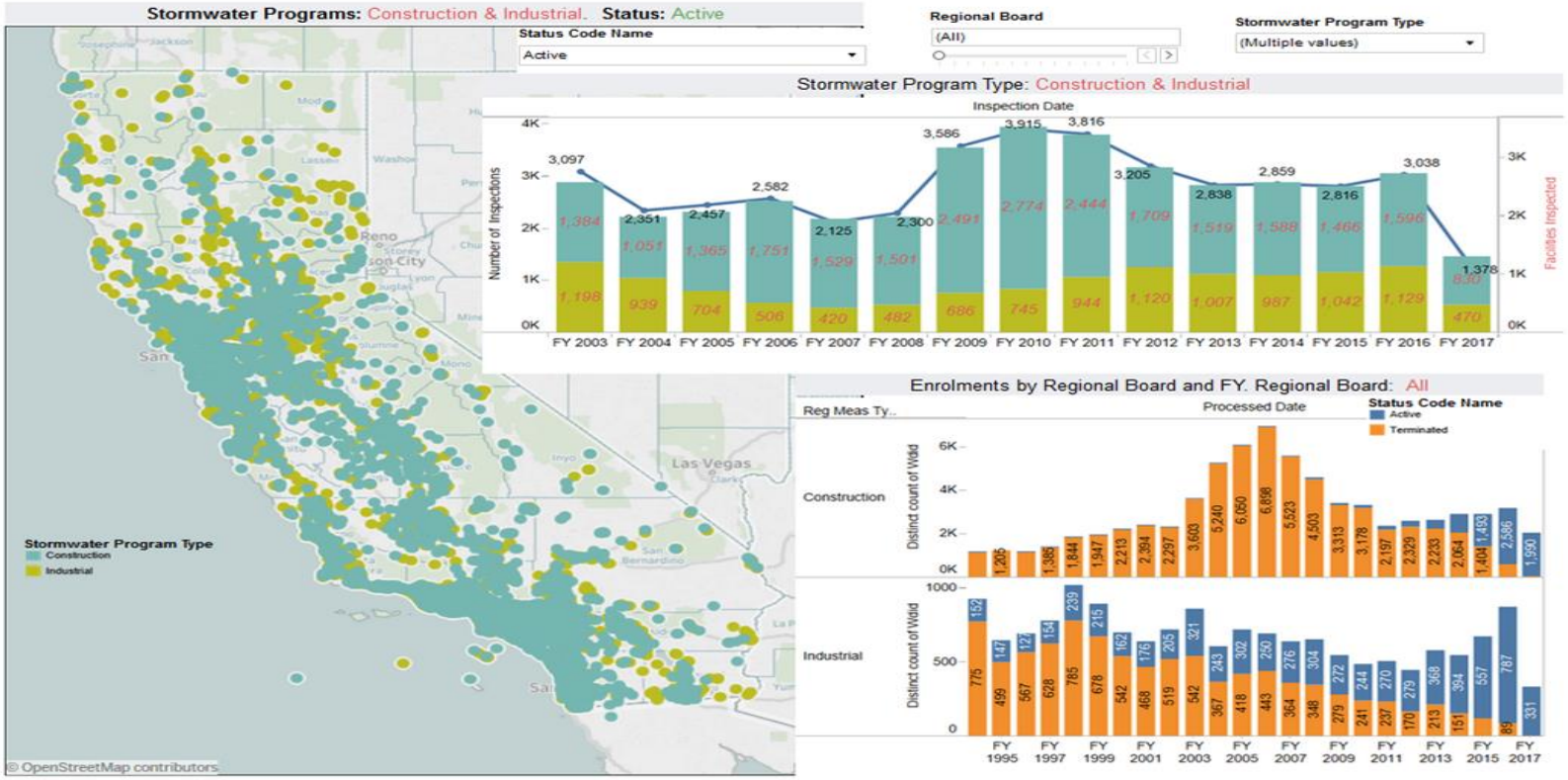
Data Indicators

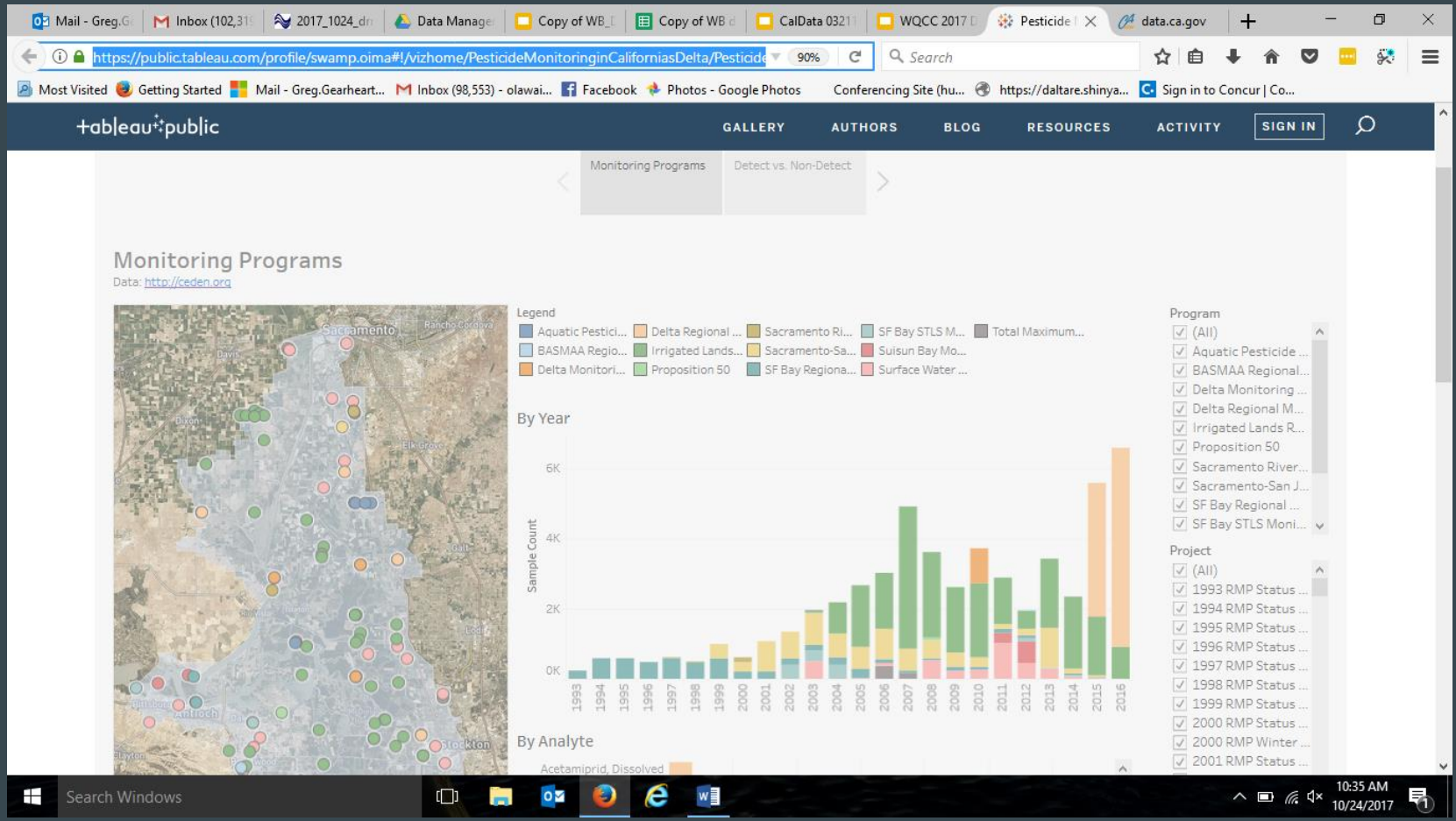
Establishing flow targets based on ecology requires large data sets of both biological and hydrological conditions in order to derive relationships that are applicable to streams across a broad range of conditions. SWAMP's large bioassessment dataset (20 years) and tools made establishing flow targets possible. The dataset, collected by

We build data visualization tools as part of the performance management program but also for “clients” and important questions that come up with the Water Boards work. Our aim is to demonstrate this capacity to other staff throughout the organization and them build this capacity so we are not the only ones doing this critical data--> information work.

Data Visualizations and Dashboards

Storm Water Program: Inspections and | Stormwater Inspections by Inspector Name | Stormwater Facilities: Active but Never Inspected | Stormwater Facility Location





https://public.tableau.com/profile/swamp.oima#!/vizhome/PesticideMonitoringinCaliforniasDelta/PesticidesStory

We regularly engage the water data community in California with our data, and use this to learn more about their interests, needs and prioritize data management infrastructure work.

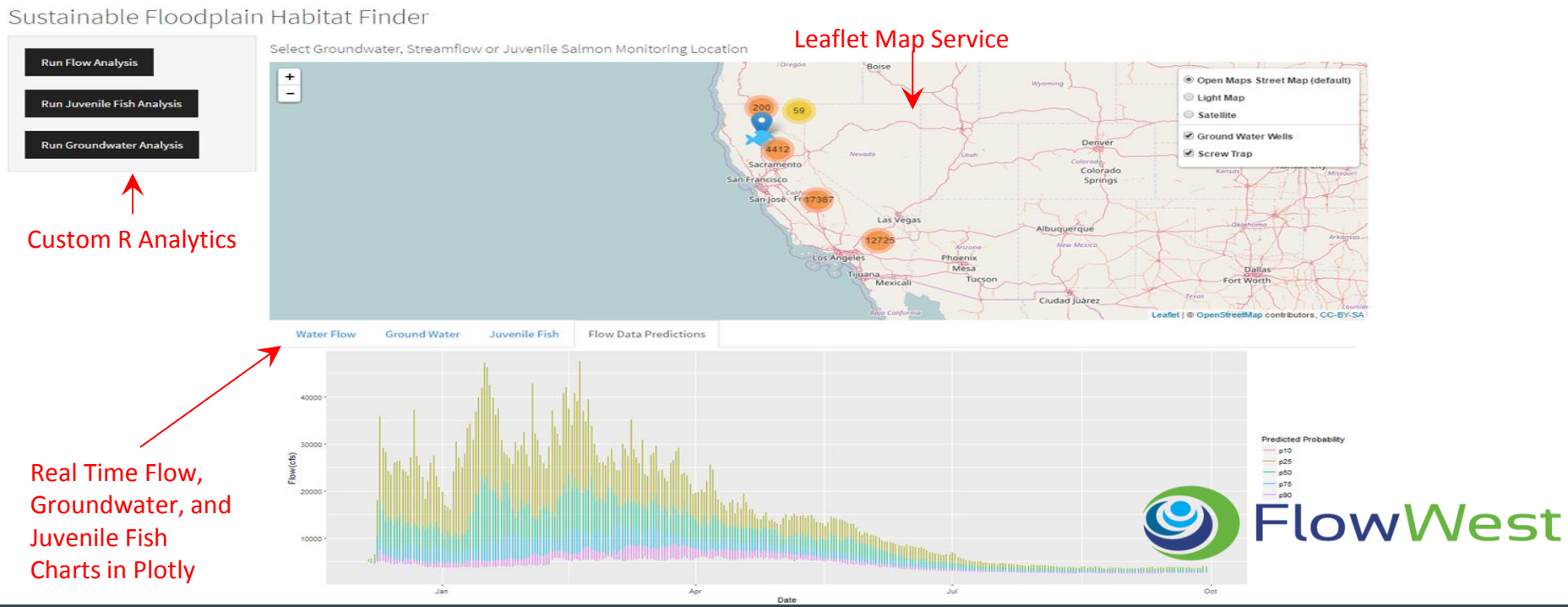
Civic Engagement Events

Water Data Civic Engagement Events 2017

- March 30, 2017 - Data Fair and Information Summit on NPDES Program
- May 2017 (not set yet) - Possible Open Data Challenge with ImagineH2O, Governor's Office and DWR
- June 2017 - 3rd Annual Watershed Health Indicators and Data Science Symposium
- https://www.waterboards.ca.gov/resources/data_databases/

Sustainable Floodplain Habitat Finder

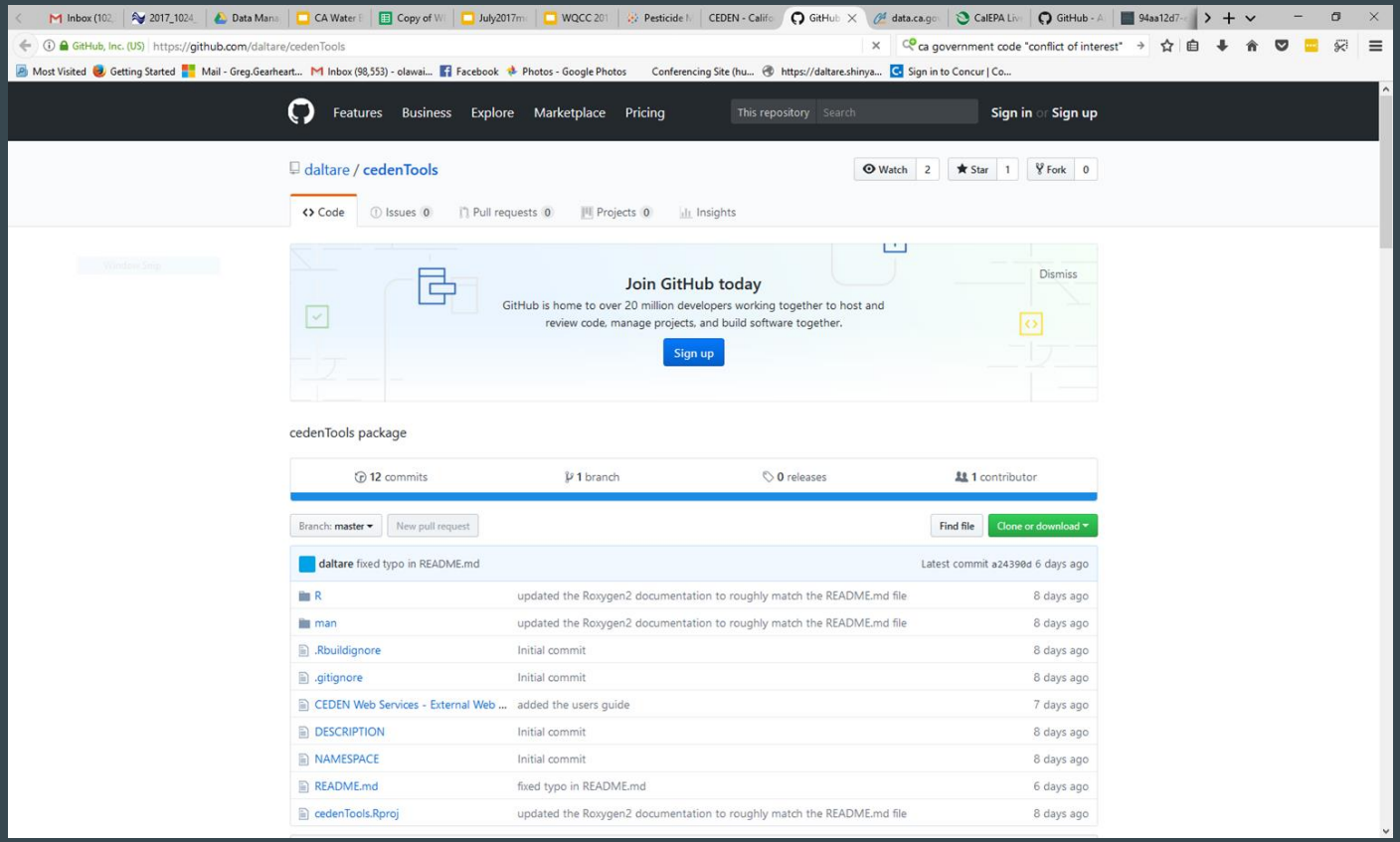
Sustainable Floodplain Habitat Finder



We build data science tools, like scripts in R and python to make our data more accessible, and explore use of artificial intelligence (AI) to to collect, analyze and generate new data from remote sensing technologies and images.

Data Science and Innovations

R package to use web services to access data in CEDEN



We have proof of concept results using computer vision (a form of artificial intelligence and machine learning) to recognize trash shapes in images, which can be captured via street sweepers, refuse trucks or other means.

There is pleasure in the pathless woods, there is rapture in the lonely shore, there is society where none intrudes, by the deep sea, and music in its roar; I love not Man the less, but Nature more. Lord Byron

greg.gearheart@waterboards.ca.gov

Greg Gearheart
Deputy Director, Office of Information Management and Analysis
CA Water Boards / CalEPA / P: 916.341.5892

WB Open Data Platform (category “water”): http://data.ca.gov/
Twitter: https://twitter.com/CaWaterDataDive